

生态环境与统计模型

2013全国大学生统计建模大赛获奖论文选

全国大学生统计建模大赛执行委员会
国家统计局统计教育培训中心

编

 中国统计出版社
China Statistics Press

生态环境与统计模型

2013全国大学生统计建模大赛获奖论文选

全国大学生统计建模大赛执行委员会
国家统计局统计教育培训中心 编



中国统计出版社
China Statistics Press

图书在版编目(CIP)数据

生态环境与统计模型：2013 全国大学生统计建模大赛获奖论文选 / 全国大学生统计建模大赛执行委员会，国家统计局统计教育培训中心编. —北京：中国统计出版社，2014. 5

ISBN 978-7-5037-7085-2

I. ①生… II. ①全… ②国… III. ①统计模型—文集 IV. ①C8-53

中国版本图书馆 CIP 数据核字 (2014) 第 093012 号

生态环境与统计模型：2013 全国大学生统计建模大赛获奖论文选

作 者/全国大学生统计建模大赛执行委员会 国家统计局统计教育培训中心

责任编辑/张 赏

特约编辑/孙 慧 李 锐

封面设计/李雪燕

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/<http://csp.stats.gov.cn/>

印 刷/河北天普润印刷厂

经 销/新华书店

开 本/710×1000mm 1/16

字 数/660 千字

印 张/39.5

版 别/2014 年 5 月第 1 版

版 次/2014 年 5 月第 1 次印刷

定 价/66.00 元

版权所有。未经许可，本书的任何部分不得以任何方式在世界任何地区以任何文字翻印、拷贝、仿制或转载。如有印装差错，由本社发行部调换。

编者说明

党的十八大提出:大力推进生态文明建设。建设生态文明,是关系人民福祉,关乎民族未来的长远大计。面对资源约束趋紧、环境污染严重、生态系统退化的严峻形势,必须树立尊重自然、顺应自然、保护自然的生态文明理念,把生态文明建设放在突出地位。

2013年全国大学生统计建模大赛以“环境”为主题,倡导生态文明理念,论文选题涉及环境与经济、健康、疾病、交通、城市化、能源、气候等诸多方面问题,选题角度多样,内容深入,切合实际。本次建模大赛由国家统计局统计教育培训中心和中国统计教育学会联合主办,是继2009年、2011年之后的第三届全国大学生统计建模比赛。为了确保大赛的公平性、公开性和公正性,各参赛队的报名以学校为单位,由各有关高等院校本科生、硕士研究生自由组队参赛,共有来自103所高等院校的508支代表队报名参赛。大赛于2013年4月发出通知,6月底之前提交参赛作品。2013年9月,大赛执委会选拔出18支参赛队进京参加决赛。经过两天紧张激烈的决赛答辩,2013年全国大学生统计建模大赛结果于9月8日下午揭晓,并在国家统计局举行了隆重的颁奖典礼。

国家统计局党组成员、总工程师郑京平出席颁奖仪式并致辞。他指出:统计建模对于大学生们形成良好的统计思维,更好地把信息技术应用到统计分析研究中去,具有着十分重要的意义。同学们通过统计建模实践,亲身参加了将理论知识应用于实践的尝试,亲身参加了发现和创造的过程,取得了在课堂里和书本上无法获得的宝贵经验和感受。希望这种以统计应用实践为主要内容的教研活动能不断得到拓展。

本届大赛评选出获奖论文124篇,其中,一等奖:本科生组3篇、研究生组3篇,二等奖:本科生组5篇、研究生组5篇,三等奖:本科生组13篇、研究生组12篇,优秀奖:本科生组45篇、研究生组38篇。中国统计教育学会网站<http://www.sescn.org.cn>开设了全国大学生统计建模专题网页,内容包括统计建模相关信息以及历届统计建模获奖论文。本书编选了此次大赛的部分获奖作品,并且请有关参赛队员根据专家提出的意见逐一进行了修改和完善。尽管如此,作品中也难以避免有错漏不当之处,还望各位专家和读者批评指正。

全国大学生统计建模大赛执委会

2013.9

目 录

一、2013 全国大学生统计建模大赛一等奖论文

本科生组

1. 基于贝叶斯时空统计方法探讨环境因素对全国手足口病的影响效应
首都医科大学 王超、刘青青、曹凯(3)
2. 基于蒙特卡罗模拟及地理信息系统的城市内涝灾害风险分析
吉林财经大学 董欣焱、黄妙琪、刘雨杭(21)
3. 禽流感病毒传播问题研究
华东师范大学 朱晓雨、吕慧娴、罗杨(57)

研究生组

1. 广州市区空气环境的 NO₂ 浓度空间分布估计
暨南大学 常启辉、方媛、吴敏(92)
2. 考虑环境污染的中国居民健康状况影响因素研究
东北财经大学 赵楠、尚继霞、张冰(117)
3. 城市化与多层次生态环境的交互耦合机制研究及预测分析
中国药科大学 陶禹希、刘昊晨、杨森(133)

二、2013 全国大学生统计建模大赛二等奖论文

本科生组

1. 应用广义估计方程分析环境与气象因素对我国手足口病发病人数的影响
南方医科大学 沈双全、刘薇、占步祥(163)
2. 空气质量与汽车保有量关系的模型
广东外语外贸大学 陈宜吟、李建兰、王玲婕(184)
3. 南京市居民生活环境满意度的多元回归模型构建与分析
南京人口管理干部学院 杨阳、单洁、孙海艳(226)
4. 基于 PLS-SEM 的药品安全环境满意度统计评估
浙江财经大学 王佳、李毓君、苏坦(254)
5. 基于相对熵组合赋权的水资源短缺风险综合评价模型
安徽大学 梁雪芹、黄婷婷、陈楼(290)

研究生组

1. 基于 Bayes 时空交互模型探讨环境因素对北京市结核病流行趋势影响规律

- 首都医科大学 潘蕾、陶丽新、王晶晶(327)
- 2.我国区域呼吸道疾病与大气环境的关联研究
中南财经政法大学 吴文娟、吴悦、岳靓(346)
- 3.河南省痢疾发病数与环境关系的统计分析建模
山东大学 高强、付金龙、丁国永(378)
- 4.基于大气环境区域性污染物排放总量分析对"联防联控"体制实施的建议
四川大学 王雅潞、曾怡、余加奎(418)
- 5.PM_{2.5}浓度的环境影响因素及其多元回归预测模型
同济大学 周游、杨潇、张静宵(456)

三、2013 全国大学生统计建模大赛三等奖论文(部分选编)

本科生组

- 1.我国城镇居民推行阶梯电价的实施现状和环境价值分析
江西财经大学 王政、刘阳、江德浩(485)
- 2.青岛市表层土壤中重金属的富集、分布特征与环境效应
华东师范大学 张晓民、孙姝婷、刘映(515)

研究生组

- 1.基于生命周期评价(LCA)技术的工业产品的环境影响分析
山东大学 龚宝龙、谭玉珍、李天宇(567)
- 2.露天烧烤对环境影响的研究
东北师范大学 李海峰、袁君君、郝姣姣(587)
- 3.环境空气质量预测及气候衍生品定价模型研究与应用
浙江财经大学 李宗龙、谷佳音、王冰灿(608)

一、2013 全国大学生统计建模大赛
一等奖论文

基于贝叶斯时空统计方法探讨环境因素 对全国手足口病的影响效应

首都医科大学 王超 刘青青 曹凯

摘 要

手足口病(Hand-foot-mouth disease, HFMD)是肠道病毒引起的常见传染病之一,可引起发热和手足、口腔等部位的皮疹、溃疡,少数患者可并发脑炎、呼吸道感染和心肌炎等,个别重症患儿病情进展快,易发生死亡。HFMD的发生与流行是病原体、环境及宿主之间相互作用的结果,其影响因素较多,并且在时空传播上具有多尺度效应。虽然我国已于2008年将HFMD纳入法定传染病,建立起较完备的流行病学监测网,积累了大量的手足口病疫情方面的信息数据,但大多数资料是孤立、分散的,技术上也难以实现集成分析与联合处理,极大地限制了其在指导制定预防措施方面发挥的作用。充分发挥HFMD疫情信息数据在指导和制定预防措施方面的作用,首先必须深刻了解HFMD发病的季节特点和地区差异,剖析影响其时空特征和生态学因素,在调整混杂因素后才能较为准确地描述其发病特征,为制定预防控制措施提供依据。

本文使用全国340个地区2008—2010年手足口病报告病例数据和全国304个气象监测站各月份气象监测指标,在贝叶斯统计思想下采用空间异质模型和空间相关模型为基础建立6个具体模型,以模型的DIC值最小为标准找出最优模型,分析平均气温、平均气压、日照时数、平均相对湿度、平均降雨量、平均日降雨量 $\geq 0.1\text{mm}$ 的日数、平均风速等气象因素对手足口病患病的影晌,以期从时空演化特征和生态影响因素方面动态分析手足口病的流行特征。分析结果显示,时空交互的CH+UH模型拟合效果最优(DIC=35507.2),HFMD的发生与时间和空间的交互作用有关,同时受到降雨量、平均气压、平均温度、平均相对湿度和日照时数等气象环境因素的影响。

1 研究背景

1.1 手足口病时空研究现况

手足口病(Hand-foot-mouth disease, HFMD)是肠道病毒引起的常见传

染病之一,其临床症状有发热和手足、口腔等部位的皮疹、溃疡,也有并发脱甲的报道^[1],少数患者可并发无菌性脑膜炎、脑炎、急性弛缓性麻痹、呼吸道感染和心肌炎等^[2],个别重症患儿病情进展快,易发生死亡^[3]。手足口病的发生与流行是病原体、环境及宿主之间相互作用的结果,其影响因素较多并且在时空传播上具有多尺度效应^[4]。我国卫生部于2008年5月2日将手足口病纳入法定丙类传染病,但属乙类管理,需要报卡登记。虽然目前已建立起较完备的流行病学监测网,积累了大量的手足口病疫情方面的信息数据,但大多数资料是孤立、分散的,技术上也难以实现集成分析与联合处理,极大地限制了其在指导制定预防措施方面的作用。因此,研究手足口病的不同时空下的患病危险度并探讨影响手足口病的环境因子,对于制定手足口病的防控措施有着十分重要的意义。

对手足口病空间流行特性的研究大多只限于空间分布的描述。随着空间分析软件和3S(GPS, GIS和RS)技术的发展^[5],动态分析传染病的时间与空间分布特征,从全新的角度来研究和认识传染病成为可能。国内学者对手足口病空间聚集性、发病率随时间变化^[6,7]进行了大量研究,对手足口病传播类型及发病危险因素^[8]也进行了一定的探索。张文增等^[9]采用Flexible空间扫描统计量对2009年1月—2010年12月北京市顺义区手足口病发病进行空间聚集性分析,发现在手足口病聚集性的时间段内其空间聚集区域存在明显差异。Wang等^[10]用贝叶斯最大熵(BME)方法和自组织映射(SOM)算法探讨了手足口病时空传播与气象因素的关系,发现HFMD病例聚集与相应地区月降雨类型相关。Deng等^[11]用Kulldorff扫描统计方法研究了广东省手足口病的时空聚集性和危险因素,发现月平均气温、相对湿度和日照时数是手足口病的危险因素。Liu等^[12]用全局自相关统计量Moran's I指数分析了山东省2007—2011年手足口发病的时空聚集性,共发现了7个时空聚集区域。Huang^[13]和Hu^[14]等用地理加权回归分别分析了广州市2008—2011年和全国2912县2008年5月的手足口发病与气象因素的关系,均发现手足口病的发生与气象因素相关,Huang还利用时间序列的方法发现手足口病的发生与前一周的气温和相对湿度具有重要联系。

上述研究发现手足口发病具有一定的空间聚集性,其发病率水平和发病率下降坡度随时间变化,而且气象因素等是手足口发病的危险因素,但其研究内容大多仅单纯描述了手足口病的时间和空间分布,分析了气象环境因素与手足口病的相关性,缺乏对疾病发病的时空交互作用的更深层次的探讨。本研究旨在把贝叶斯统计方法应用到手足口病流行病学分析中,以期分析时空交互作用对手足口病时空演化特征的影响,动态分析气象环境因素对全国手足口病的影响。

1.2 贝叶斯统计理论与贝叶斯时空模型

贝叶斯统计是基于总体信息、样本信息和先验信息(即在抽样之前有关统计问题的一些信息,主要来源于经验和历史资料)进行统计推断的统计学理论。贝

叶斯法则指当分析样本大到接近总体数时,样本中事件发生的概率将接近于总体中事件发生的概率。贝叶斯统计认为:(1)未知参数是随机变量而不是常数。人们在获得未知参数前,对它的取值具有事先的主观认识或分析,这种认识可用概率分布表示,称为先验分布。这是贝叶斯学派最基本的观点^[15]。(2)进行统计推断时,除了依据总体信息和样本信息外还利用先验信息,用概率论中求条件概率分布方法,可以得出后验概率。贝叶斯公式为:

$$P(A | B) = \frac{P(B | A) P(A)}{P(B)} \propto L(A | B) P(A)$$

其中 $L(A|B)$ 是在 B 发生的情况下 A 发生的可能性。 $P(A)$ 是 A 的先验概率,之所以称为“先验”是因为它不考虑任何 B 方面的因素。 $P(A|B)$ 是已知 B 发生后 A 的条件概率,也由于得自 B 的取值而被称作 A 的后验概率。 $P(B|A)$ 是已知 A 发生后 B 的条件概率,也由于得自 A 的取值而被称作 B 的后验概率。 $P(B)$ 是 B 的先验概率。

后验分布是用抽样信息对先验分布作调整的结果,通过求解后验分布的特征如后验均数、中位数、方差和百分位数等即可对未知参数做出推断,这是贝叶斯学派进行统计推断的基础^[16]。由于对后验分布复杂的计算的问题,贝叶斯统计并没有像经典频率统计那样得到广泛认可与应用。近年来随着 MCMC(Monte Carlo Markov Chain)的发展以及贝叶斯统计软件 WinBUGS 等的推广,贝叶斯统计开始在各个方面运用发展起来。

贝叶斯时空模型是在贝叶斯统计思想(即结合样本信息与先验分布对后验分布参数做出估计)的框架下,为分析时空数据中蕴含的时间和空间信息而建立起来的数学模型^[17]。与传统的空间流行病学研究方法相比,将时间与空间结合起来分析,在具有空间相关关系疾病的研究中能更加准确的反映疾病发生流行的状态^[18]。Lawson^[19]提出了用贝叶斯时空 SIR 模型分析美国 Carolina 南部各个城镇 2004—2005 年流感流行季节周统计的流感疫情,证实了把贝叶斯统计方法融入传染病时空传播模型分析中,更有助于深入了解传染病在人群中的时空传播机制和规律,改变和发现对传染病发生、发展及消亡有影响的各种因素,科学制定和调整防控策略。贝叶斯时空模型认为模型中的所有未知参数为随机变量,需要利用先验分布进行描述,发病或死亡危险度在空间和时间上的变异用随机效应来表示。

1.3 生态学分析方法

近年来,对于手足口病患病的危险因素的研究有基于个体水平的危险因素(如个人卫生习惯、临床症状特征等)研究和基于群体水平的危险因素(人口密度、气象环境因素等)研究。与基于个体水平的研究相比,生态学分析技术(Ecological Analysis)能够从整体水平上研究相关因素与疾病发生的关系,避免了由于个体之

间不同因素带来的偏倚。生态学分析用于疾病结局变量中具有地理信息并且疾病发生与协变量或预测因素有关的数据分析^[20]。这些协变量可以在空间聚集的不同水平上获得。比如我们希望研究疾病的发生是否与月平均温度有关,我们既要得到每个区域的疾病发生数 y_i ,也要得到该地区月平均温度。其基本模型为: $y_i \sim f(\mu_i)$, 其中, $E(y_i) = \mu_i$, $g(\mu_i) = S(x_{1i})$, $g(\mu_i)$ 是连接函数, $S(x_{1i})$ 是协变量的线性或非线性函数。最简单的模型为:

$$y_i \sim \text{Pois}(\mu_i), \mu_i = e_i \theta_i = e_i \exp\{\alpha_0 + \alpha_1 x_{1i}\},$$

其中, $\exp\{\alpha_0\}$ 是全局缩放参数, $\exp\{\alpha_1 x_{1i}\}$ 是协变量的调整。统计推断最关心的是 θ , α_0 和 α_1 。

综上我们推断,使用贝叶斯时空模型(尤其是加入时空交互效应后的模型)与生态学分析能够更准确地探讨气象环境因素在群体水平上对手足口病发病的影响。

2 数据描述

2.1 手足口病数据信息

本研究所用的手足口病信息来自从中国疾病预防控制中心传染病网络直报系统中提取的 2008—2010 年手足口病疫情数据病例信息,包含全国各地从 2008 年 1 月—2010 年 12 月登记的所有手足口病数据(包括各地区发病人数、发病率等)。经过整理、排除建设兵团等发病地区未知的发病数据后最终纳入 340 个地区的 3396797 例手足口病报告数据作为研究对象。

2.2 全国各地数字区划图

全国各地数字区划地图来自于北京超图软件有限公司,比例为 1:25000000,格式为 ArcGis 平台 .shp 文件,地图文件中把全国分为 347 个地区。在进行空间分析时,因为地图中的香港、澳门、台湾、东沙群岛、伊犁地区、石河子市没有相应的发病数据,因此在制图分析时,未将这些地区纳入;重庆市的发病数据有重庆市辖区和重庆郊区县两部分,而地图中还有重庆郊区市,为了使两者保持一致,在制图分析时将重庆郊区县和重庆郊区市合并为重庆郊区县进行分析。

2.3 环境数据

气象环境数据来自中国气象科学数据共享服务网(China Meteorological Data Sharing Service System),数据包括全国 674 个气象监测站各月份气象监测指标。本次研究选用的气象资料筛选过程如下:将各监测站经纬度标注到数字区划地图中,把与各地区区域中心的距离最近的气象站的监测数据作为相应地区的气象数据,整理后共有 304 个气象监测点数据纳入研究。由于月份气象监测指标过多,在参考了大量文献后,同时为了保证模型拟合效果,所以决定分析的变量有降水量(mm)、平均监测站气压(hPa)、平均风速(m/s)、平均气温($^{\circ}\text{C}$)、平均相对湿度(%),日降水量 $\geq 0.1\text{mm}$ 天数(天)和日照时数(小时)7 个气象环境指标。北京

北京市市辖区 2008—2010 年各月份手足口病报告病例数与气象环境因素(平均气压除外)变化曲线如图 1 所示。

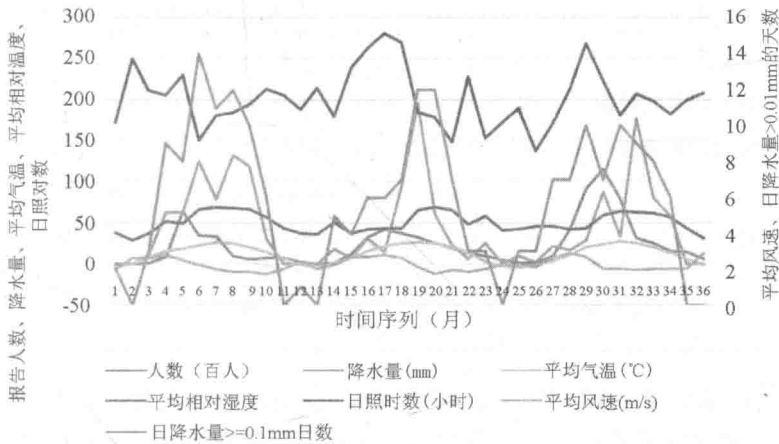


图 1 北京市市辖区手足口病报告病例数与气象环境因素的时间变化曲线

在初始分析中,时间按月共分为 36 组进行分析,但由于数据量太大,模型难以收敛,最终依据数据特点将相邻三个月进行合并,将 36 个月份划分成 12 组,重新整理发病与气象环境资料得到研究所用数据库。

3 模型建立与求解

3.1 Bayes 时空模型的建立

Bayes 时空模型的基本思想是基于 Bayes 理论,即结合样本信息与先验分布对后验分布参数做出估计,可以估计其后验特征,包括均数,中位数,方差,百分位数等。泊松模型是小区域数据最常用的模型,适用于疾病发生数较小和人口基数较大的小区域。通常 y_i 表示疾病发生例数,服从期望为 μ_i 的 Poisson 分布,其似然函数为:

$$L(y | \mu) = \prod_{i=1}^m \mu_i^{y_i} \exp(-\mu_i) / y_i!$$

均值函数由两部分组成:(1)背景人口效应;(2)每个区域的超额危险度,也常被叫做相对危险度。第一部分通常是由标准化的人口疾病发生率和局部期望发生率的比较估计得到的。因此,假设数据满足独立分布,则:

$$E(y_i) = \mu_i = e_i \theta_i,$$

其中 e_i 是第 i 个区域的期望发病人数, θ_i 是第 i 个区域的相对危险度,即为研究所关心的指标,同时考虑时间维度与疾病发生的关系,建立模型如下,6 种模型见表 1。

$$y_{ij} \sim Pois(\mu_{ij}), \mu_{ij} = e_{ij}\theta_{ij}$$

表 1 Bayes 泊松模型的建立

模型	相对危险度估计
UH 模型	$\theta_{ij} = \exp(\alpha_0 + \alpha_{1j} + v_i)$
UH 模型+自回归时间效应	$\theta_{ij} = \exp(\alpha_0 + \alpha_{1j} + v_i + g_j)$
UH 模型+时间趋势效应	$\theta_{ij} = \exp(\alpha_0 + \alpha_{1t_j} + v_i + g_j)$
CH 模型	$\theta_{ij} = \exp(\alpha_0 + \alpha_1 * time1_j + u_i + g_j + d_i * time1_j)$
CH 模型+ UH 模型	$\theta_{ij} = \exp(\alpha_0 + \alpha_1 * time1_j + u_i + v_i + g_j + d_i * time1_j)$
CH+UH 模型+时空交互模型	$\theta_{ij} = \exp(\alpha_0 + \alpha_1 * time1_j + u_i + v_i + g_j + \psi_{ij} + d_i * time1_j)$

表 1 中 i 表示第 i 个区县, j 表示第 j 年, y_{ij} 表示患病人数, e_{ij} 表示期望发病人数, θ_{ij} 表示患病危险度, α_0 是截距, α_{1j} 是第 j 年的截距, v_i 表示不相关空间异质效应, 即由非空间因素引起的随机效应。 u_i 表示相关空间异质效应, 即由空间因素引起的随机效应。 g_j 表示自回归时间效应, t_j 表示时间趋势效应, α_1 表示时间效应的系数, $time1_j$ 表示第 j 年的时间效应, d_i 表示距离效应, ψ_{ij} 表示时空交互效应。

以上模型均在 WinBUGS1. 4. 3 中实现, 经过模型的加工 (Specification)、更新 (Update) 进而推断样本参数 (Samples), 得到后验分布的均值、众数、标准差和可信区间 (Credible Interval) 等。 此过程采用蒙特卡洛马尔可夫算法 (MCMC)。

贝叶斯模型中的后验分布是在综合了样本信息和先验分布后经过模型迭代抽样以后得到的, 受先验分布的影响较小^[21]。 本次研究先验分布设置如下: UH (Uncorrelated Heterogeneity Models) 模型, 即空间不相关异质模型, 该模型认为待研究的地区之间不存在空间上的相关关系, 某一地区的协变量只影响相应地区结局变量, 在模型中用 v_i 表示, 其先验分布选择正态分布。 CH (Correlated Heterogeneity Models) 模型, 即空间相关异质模型, 模型认为某一地区的结局变量不仅受到本地区的协变量的影响, 还可能受到邻近地区的影响, 相互临近的地区不能单独分析, 在模型中用 u_i 表示。 其先验分布选择现在最为广泛使用的 CAR 模型^[21] (conditional autoregressive model, 即条件自回归模型)。 地区间的空间关系用简单邻接权重矩阵 (simple adjacency weights matrix) 表示, 如果两个区县相邻, 则权重赋值为 1; 反之, 赋值为 0。 截距项的先验分布因未找到相关资料而选择无信息先验分布 $dflat()$, 模型参数的先验分布选择正态分布 $dnorm(0, \tau\alpha\beta)$, 其参数的先验分布即超先验分布选择均匀分布 $U(0, 10)$ 。

3.2 气象环境因素分析

生态学分析用于疾病结局变量具有地理信息, 并且疾病发生与协变量或预测

因素有关的数据分析。这些协变量可以在空间聚集的不同水平上获得。本研究将第一部分中的最优模型时空交互 CH+UH 模型带入协变量:降水量(mm)、平均监测站气压(hPa)、平均风速(m/s)、平均气温(℃)、平均相对湿度(%)、日降水量 $\geq 0.1\text{mm}$ 天数(天)和日照时数(小时)。模型如下及图 2 所示:

$$y_{ij} = \text{Pois}(\mu_{ij})$$

$$\mu_{ij} = e_{ij} \theta_{ij}$$

$$\theta_{ij} = \exp(\alpha_0 + \alpha_1 * \text{time}_{ij} + \mu_i + g_i + \text{psi}_{ij} + \beta_1 * \text{rf}_{ij} + \beta_2 * \text{aap}_{ij} + \beta_3 * \text{aws}_{ij} + \beta_4 * \text{aat}_{ij} + \beta_5 * \text{arh}_{ij} + \beta_6 * \text{dr}_{ij} + \beta_7 * \text{st}_{ij})$$

式中, $\beta_1 - \beta_7$ 表示协变量的回归系数。

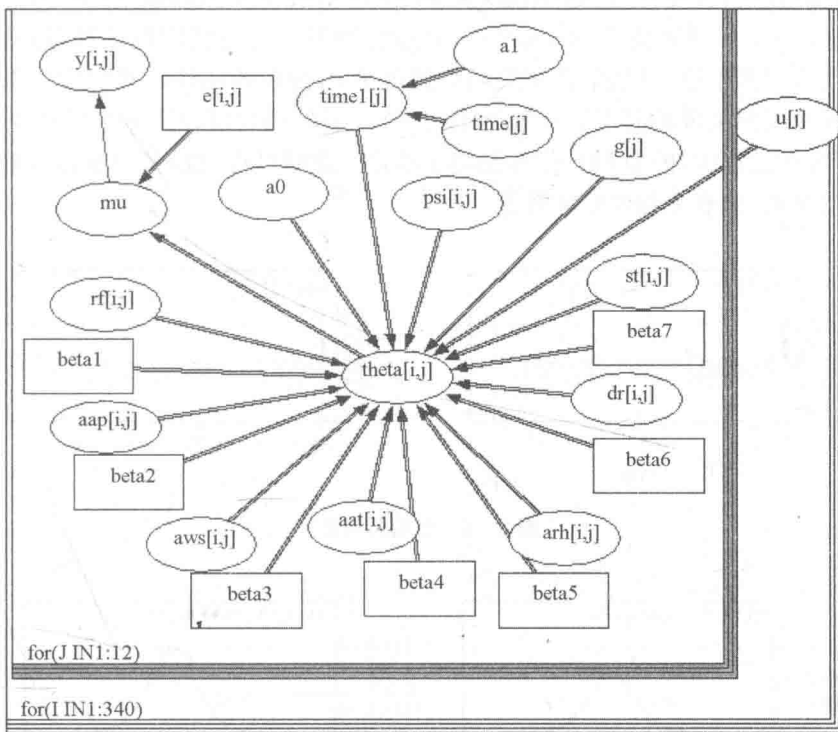


图 2 有向图模型结构示意图

3.3 模型求解

对于任何以 MCMC 为基础的完全概率模型分析,关于 MCMC 模拟的监测和收敛性的判断都是非常重要的一步。一个 MCMC 模拟称为收敛的,是指模拟结果来自于真的 Markov Chain 的稳定或目标分布。能从理论上证明收敛性的情形非常少,大多数情况下,我们只能满足于有效收敛(effective convergence)。

本次研究的模型求解在 WinBUGS1.4.3 软件中实现。在 WinBUGS 软件主

窗口下先后载入模型代码、分析所用协变量数据和模型中应用的空间矩阵(分析空间相关模型时)的数据,在 SpecificationTool 工具条中将上述数据合并,载入模型的初始值(部分初始值为软件自动生成);在 Sample Monitor Tool 工具条中设置本次研究关心的参数(模型中各协变量的系数、指代不同时间空间效应的指标等);然后在 Updata Tool 中设置迭代更新的次数,开始进行模型的迭代拟合。模型迭代每完成一次时,查看模型的拟合效果。

通过 Gibbs 抽样,反复迭代,约 45000 次后,为了降低初始值的影响,舍去(burn in)前 5000 次的资料,由抽样动态轨迹图(dynamic trace plot)显示达到基本平稳,各参数估计值趋于稳定,迭代过程基本收敛。此时模型迭代收敛性的检验可以通过迭代轨迹(trace)、迭代历史(history)、自相关函数(auto cor)、分位数估计(quantiles)、核密度图(density)等进行判断。当迭代次数足够多,迭代轨迹、迭代历史基本趋于稳定,自相关函数接近于 0,分位数估计图近似为三条水平直线,核密度图呈现类似正态分布图(图 3-1—图 3-5),可以认为迭代已经收敛,在模型收敛过程中,模型 DIC 值逐渐趋于稳定。此时可得到模型中参数后验分布的均值、众数、标准差和可信区间等。

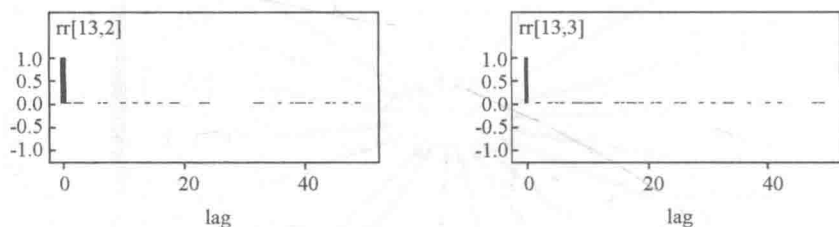


图 3-1 自相关函数

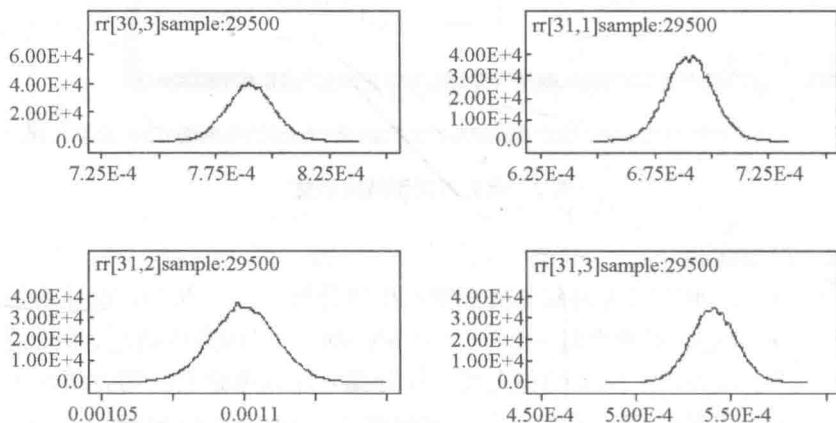


图 3-2 核密度图

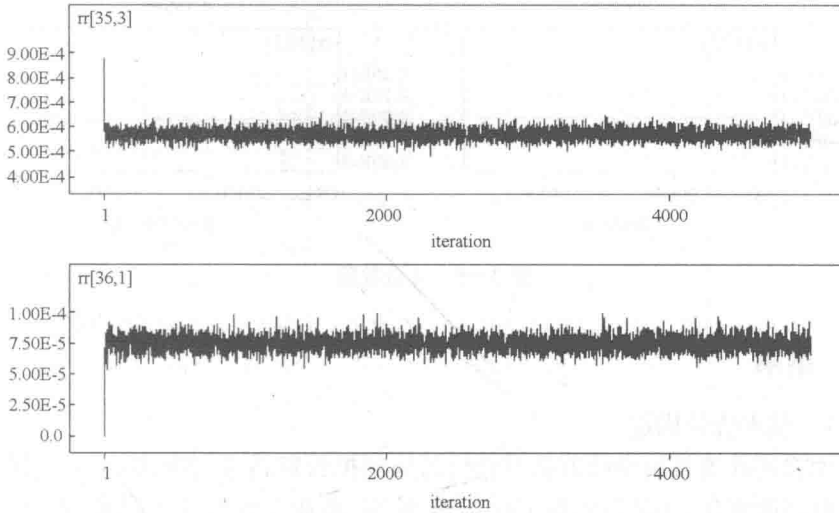


图 3-3 迭代历史图

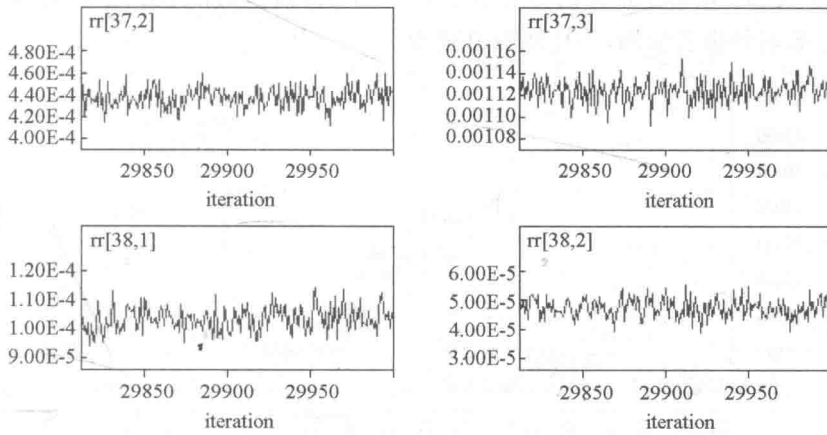


图 3-4 迭代轨迹图

